

VISIT...

LANZAROTE
Caliente.COM

15

***E. coli* Expression Systems**

Peter A. Bell

Expression Vector Structure	462
What Makes a Plasmid an Expression Vector?	462
Is a Stronger Promoter Always Desirable?	463
Why Do Promoters Leak and What Can You Do about It?	464
What Factors Affect the Level of Translation?	464
What Can Affect the Stability of the Protein in the Cell?	464
Which Protein Expression System Suits Your Needs?	465
Track Record	465
What Do You Know about the Gene to Be Expressed? ...	465
What Do You Know about Your Protein?	468
Advertisements for Commercial Expression Vectors Are Very Promising. What Levels of Expression Should You Expect?	470
Which <i>E. coli</i> Strain Will Provide Maximal Expression for Your Clone?	471
Why Should You Select a Fusion System?	471
When Should You Avoid a Fusion System?	472
Is It Necessary to Cleave the Tag off the Fusion Protein?	474
Will Extra Amino Acid Residues Affect Your Protein of Interest after Digestion?	475
Working with Expression Systems	475
What Are the Options for Cloning a Gene for Expression?	475

Is Screening Necessary Prior to Expression?	476
What Aspects of Growth and Induction Are Critical to Success?	478
What Are the Options for Lysing Cells?	479
Troubleshooting	480
No Expression of the Protein	480
The Protein Is Expressed, but It Is Not the Expected Size Based on Electrophoretic Analysis	480
The Protein Is Insoluble, Now What?	481
Solubility Is Essential. What Are Your Options?	482
The Protein Is Made, but Very Little Is Full-Length; Most of It Is Cleaved to Smaller Fragments	483
Your Fusion Protein Won't Bind to Its Affinity Resin	484
Your Fusion Protein Won't Digest	485
Cleavage of the Fusion Protein with a Protease Produced Several Extra Bands	485
Extra Protein Bands Are Observed after Affinity Purification	485
Must the Protease Be Removed after Digestion of the Fusion Protein?	486
Bibliography	486

Over the past decade the variety of hosts and vector systems for recombinant protein expression has increased dramatically. Researchers now select from among mammalian, insect, yeast, and prokaryotic hosts, and the number of vectors available for use in these organisms continues to grow. With the increased availability of cDNAs and protein coding sequencing information, it is certain that these and other, yet to be developed systems will be important in the future. Despite the development of eukaryotic systems, *E. coli* remains the most widely used host for recombinant protein expression. *E. coli* is easy to transform, grows quickly in simple media, and requires inexpensive equipment for growth and storage. And in most cases, *E. coli* can be made to produce adequate amounts of protein suitable for the intended application. The purpose of this chapter is to guide the user in selecting the appropriate host and troubleshooting the process of recombinant protein expression.

EXPRESSION VECTOR STRUCTURE

What Makes a Plasmid an Expression Vector?

Vectors for expression in *E. coli* contain at a minimum, the following elements:

Table 15.1 Characteristics of Popular Prokaryotic Promoters

Promoter		Regulation/Inducer (Concentration)	Strength
LacUV5	Lactose operon	<i>lacI</i> /IPTG (0.1–1 mM)	Strong
Trp	Tryptophan operon	<i>trpR</i> 3- beta-indoleacrylic acid	Strong
Tac	Hybrid of –35 Trp and –10 lac promoter	<i>lacI</i> /IPTG (0.1–1 mM)	Strong
P _L	Phage lambda	Lambda <i>cI</i> repressor/heat	Strong
Phage T5	T5 phage	<i>lacI</i> (2 operators)/ IPTG (0.1–1 mM)	Strong
Arabinose	Arabinose operon	<i>AraBAD</i> /arabinose (1 μ m–10 mM)	Variable
T7	T7 phage RNA polymerase	<i>lacI</i> /IPTG (0.1–1 mM)	Very strong

- A transcriptional promoter.
- A ribosome binding site.
- A translation initiation site.
- A selective marker (e.g., antibiotic resistance).
- An origin of replication.

In general, things that affect these can affect the level of protein expression. At a minimum, transcription promoters in *E. coli* consist of two DNA hexamers located –35 and –10 relative to the transcriptional start site. Together these elements mediate binding of the about 500 kDa multimeric complex of RNA polymerase.

Suppliers of vectors for expression have selected highly active, and inducible promoter sequences, and there is usually little need to be concerned until a problem is encountered in expression. A list of the commonly used promoters and their regulation is shown in Table 15.1.

Is a Stronger Promoter Always Desirable?

A strong promoter may not be best for all situations. Overproduction of RNA may saturate translation machinery, and maximizing RNA synthesis may not be desirable or necessary. A weaker promoter may actually give higher steady-state levels of soluble, intact protein than one that is rapidly induced.

Why Do Promoters Leak and What Can You Do about It?

Most promoters will have some background activity. Promoters regulated by the lactose operator/repressor will drive a small amount of transcription in the absence of added inducer (e.g., IPTG). To minimize this leakage, 10% glucose can be added to the medium to repress the lactose induction pathway, the growth temperature can be reduced to 15 to 30°C, and a minimal medium that contains no trace amounts of lactose can be used. Promoter leakage is only a problem when the expressed protein is highly toxic to the cells.

The tightly regulated T7 promoter has very low background due to the low levels of T7 RNA polymerase made in the absence of inducer (in specifically engineered host cells such as BL21 (DE3)/pLysS). It has been estimated that the fold induction of transcription in the T7 driven pET vector system is greater than 1000, while the magnitude of induction obtained with lac repressor regulated promoters is generally about 50-fold.

What Factors Affect the Level of Translation?

Translation can be affected by nucleotides adjacent to the ATG initiator codon, the amino acid residue immediately following the initiator, and secondary structures in the vicinity of the start site. Most commercially available vectors for expression use optimal ATG and Shine-Dalgarno sequences. Secondary structures in the mRNA contributed by the gene of interest can prevent ribosome binding (Tessier et al., 1984; Looman et al., 1986; Lee et al., 1987). In addition, the downstream box AAUCACAAAGUG found after the initiator codon in many bacterial genes can also enhance translation initiation. Conversion of the amino terminal sequence of the gene of interest to one that comes close to this consensus may improve the rate of translation of the mRNA (Etchegaray and Inouye, 1999).

What Can Affect the Stability of the Protein in the Cell?

One of the first steps in protein degradation in *E. coli* is the catalyzed removal of the amino terminal methionine residue. This reaction, catalyzed by methionyl aminopeptidase, occurs more slowly when the amino acid in the +2 position has a larger side chain (Hirel et al., 1989; Lathrop et al., 1992). When the methionine residue is intact, the protein will be stable to all but endopeptidase cleavage. Tobias et al. (1991) have determined the relationship between a protein's amino terminal amino acid

and its stability in bacteria, that is, the “N-end rule.” They reported protein half-lives of only 2 minutes when the following amino acids were present at the amino terminus: Arg, Lys, Phe, Leu, Trp, and Tyr. In contrast, all other amino acids conferred half-lives of >10 hours when present at the amino terminus of the protein examined. This suggests that one should examine the sequence to be expressed for the residue in the +2 position. If the residue is among those that destabilize the protein, it may be worth the effort to change this residue to one that confers stability.

WHICH PROTEIN EXPRESSION SYSTEM SUITS YOUR NEEDS?

Track Record

What systems are currently used in the laboratory or by others in the field? If the protein coding sequence of interest is well characterized, and the protein or its close relatives have been expressed successfully by others in the field, it is wise to try the same expression system. Go with what has worked in the past. If nothing else, results obtained using the familiar system will serve as a starting point. As an example, most of the recombinant expression of mammalian src homology SH2 protein interaction domains has been done using the pGEX vector series, and similar examples of preferred systems are found in other fields of research. If little is known about the protein to be expressed, it is best to take stock of what information there is before entering the lab. Before beginning any experimentation, it is wise to answer the following question:

What Do You Know about the Gene to Be Expressed?

Source

In general, simple globular proteins from prokaryotic and eukaryotic sources are good candidates for expression in *E. coli*. Monomeric proteins with few cysteines or prosthetic groups (e.g., heme and metals) and of average size (<60kDa) will likely give good production. Secreted eukaryotic proteins and membrane-bound proteins, especially those with several transmembrane domains, are likely to be problematic in *E. coli*. Solubility of recombinant proteins in *E. coli* can also be estimated by a mathematical analysis of the amino acid sequences (Wilkinson and Harrison, 1991).

Presence of a Start Codon

Some expression vectors provide the start codon for translation initiation, while others rely on the start codon of the gene you're trying to express. Note that in *E. coli*, 5 to 12 base pairs or less separate the ribosome binding site and the start codon. So you would incorporate this requirement into your cloning strategy when the start codon is provided by the gene you plan to express.

GC Content

Coding sequences with high GC (>70%) content may reduce the level of expression of a protein in *E. coli*. Check the sequence using a DNA analysis program.

Codon Usage

Codon usage may also affect the level of protein expression. If the gene of interest contains codons not commonly used in *E. coli*, low expression may result due to the depletion of tRNAs for the rarer codons. When one or more rare codons is encountered, translational pausing may result, slowing the rate of protein synthesis and exposing the mRNA to degradation. This potential problem is of particular concern when the sequence encodes a protein >60 kDa, when rare codons are found at high frequency, or when multiple rare codons are found over a short distance of the coding sequence. For example, rare codons for arginine found in tandem can create a recognition sequence for ribosome binding (e.g., _AGGAGG) that closely approximates a Shine-Dalgarno sequence UAAGGAGG. This may bind ribosomes non-productively and block translation from the bona fide ribosome binding site (RBS) at the initiator codon further upstream. Nonetheless, the appearance of a rare codon does not necessarily lead to poor expression. It is best to try expression of the native gene, and then make changes if these seem warranted later. Strategies include mutating the gene of interest to use optimal codons for the host organism, and co-transforming the host with rare tRNA genes. In one example, introduction into the *E. coli* host of a rare arginine (AGG) tRNA resulted in a several-fold increase in the expression of a protein that uses the AGG codon (Hua et al., 1994). In another case, substitution of the rare arginine codon AGG with the *E. coli*-preferred CGU improved expression (Robinson et al., 1984). Other work has shown that rare codons account for decreased expression of the gene of interest in *E. coli* (Zhang, Zubay, and Goldman, 1991; Sorensen, Kurland, and Pederson, 1989). Rare codons may have an even more dramatic

effect on translation when they occur close to the initiator codon (Chen and Inouye, 1990). While codon usage is not the only or most important factor, be aware that it may influence translation efficiency.

Secondary Structure

Secondary structures that occur near the start codon may block translation initiation (Gold et al., 1981; Buell et al., 1985), or serve as translation pause sites resulting in premature termination and truncated protein. These can be found using DNA or RNA analysis software. Structures with clear stem structures greater than eight bases long may be disrupted by site-specific mutation or by making all or a portion of the coding sequence synthetically.

Depending on the size of the gene, and the importance of obtaining high-expression levels, it may be worth synthesizing the gene. This has been generally done by synthesizing overlapping oligonucleotides that when annealed can be extended using PCR and ligated to form the full-length coding sequence. There are several examples where this approach has been used to optimize codon usage for *E. coli* (Koshiba et al., 1999; Beck von Bodman et al., 1986). In addition, if one takes on the work and expense of synthesizing a gene, secondary structures in the predicted RNA that might stall translation can be removed, and sites for restriction endonucleases can be introduced.

Size of a Gene or Protein

As a rule, very large (>100kDa) and very small (<5kDa) proteins are more difficult to express in *E. coli*. Small polypeptides with little secondary structure tend to be rapidly degraded in *E. coli*. Degradation can be minimized by expressing such short oligopeptides as concatemers with proteolytic or chemical cleavage sites in between the monomeric units (Hostomsky, Smrt, and Paces, 1985). Short peptides are also successfully expressed as fusion proteins. Fusion with GST, MalB or other larger, well-folded partners will tend to stabilize a short peptide, making expression possible and purification relatively simple. One publication has shown MBP to be superior to other large fusion proteins at stabilizing short polypeptides (Kapust and Waugh, 1999). At the other extreme, proteins that are above 60kDa are best made using smaller affinity tags, such as FLAG, his₆, or on their own, without any fusion. While there is no clear upper limit, the larger the protein, the lower the yield is likely to be.

What Do You Know about Your Protein?

Cysteines

There are many things that *E. coli* does not do well, or at all. If the protein of interest is naturally multimeric, or requires post-translational modifications for activity, *E. coli* as an expression host may be a poor choice. Disulfide bonds, formed between two cysteines in an expressed protein, are made inefficiently in the reducing environment of the *E. coli* cytoplasm (Bessette et al., 1999; Derman et al., 1993). If the protein is produced, and can be purified from *E. coli*, in vitro oxidation of the cysteines may be tried (Dodd et al., 1995). Alternatively, the gene of interest can be cloned in a vector that includes a signal sequence (e.g., OmpA, geneIII, and phoA) that will direct the recombinant protein to the relatively oxidizing environment of the periplasm of *E. coli*, where disulfide formation is more efficient. Strains of *E. coli* that are deficient in thioredoxin reductase (trxB) permit proper disulfide formation in the cytoplasm (Derman et al., 1993; Yasukawa et al., 1995). Subsequent work has produced strains that lack both trxB and glutathione oxidoreductase and give better rates of disulfide formation than those seen in native *E. coli* periplasm (Bessette et al., 1999).

Membrane Bound

If the protein to be expressed is naturally associated with membrane and/or has at least one transmembrane domain, addition of a secretion signal to the amino terminus may help to maximize expression of functional protein. Signal sequences, about 20 residues long are derived from proteins that naturally are secreted into the periplasmic space, such as pelB, OmpA, OmpT, MalE, alkaline phosphatase (phoA), or geneIII of filamentous phage (Izard and Kendall, 1994). Protein with an amino terminal signal will be directed to the inner membrane of *E. coli*, and the carboxy terminal portion of the protein will be translocated into the periplasmic space. Depending on the hydrophobicity of the protein of interest, it may not translocate entirely into the periplasm but remain associated with the inner membrane. Secretion may help stabilize proteins from proteolytic attack (Pines and Inouye, 1999), or at least can reduce aggregation of hydrophobic proteins in the cytoplasm, and minimize inclusion body formation. Because of the reducing environment of the periplasmic space, proteins that contain one or more disulfide bonds are best secreted.

The presence of an N-terminal signal sequence appears to

be necessary but not sufficient to direct a target protein to the periplasm. Translocation across the outer membrane and into the growth medium is inefficient. In most cases target proteins found in the growth medium are the result of damage to the cell envelope and do not represent true secretion (Stader and Silhavy, 1990). Translocation across the inner cell membrane of *E. coli* is incompletely understood (reviewed by Wickner, Driessen, and Hartl, 1991), and the efficiency of export will depend on the individual target protein. Currently the export cannot be predicted based on protein sequence, although some generalizations have been made about the sequence immediately following the signal peptide (Boyd and Beckwith, 1990; Yamane and Mizushima, 1988). Therefore it is possible to find target proteins in the cytoplasm (with uncleaved signal sequence) or in the periplasm in partially processed form, in place of or in addition to the expected periplasmic processed species. In some cases the proportion of protein that is exported can be increased by lowering the temperature 15 to 30°C during induction.

Post-translational Modification

E. coli does not glycosylate or phosphorylate proteins or recognize proteolytic processing signals from eukaryotes, so take this into account when designing the cloning strategy. If proteolytic processing is needed, it is best to express only the coding sequences for the fully processed protein. If the protein of interest requires glycosylation for activity, and full activity is important in the final use, consider a eukaryotic host, such as *Pichia*, insect cells, or mammalian cells.

Is the Protein Potentially Toxic?

Consider whether the protein of interest is likely to have a toxic effect on the host cell. Where the function of the protein is known, this can be guessed at with some accuracy. For example, non-specific proteases, nucleases, or pore-forming membrane proteins might all be expected to have some toxic effect on *E. coli*. Expression of toxic proteins may be very low, and there will be strong selective pressure on cells to eliminate the gene of interest by point mutation to change the translation frame, insertion of a stop codon, or change in an amino acid residue critical to the protein's function. Larger deletion of parts of the plasmid may also be seen. If there is a suggestion that the gene product will be toxic, use an expression vector with a tightly regulated promoter (e.g., T7, pET

vectors). Minimize propagation of the cells to avoid opportunities for mutation and recombination.

Must Your Protein Be Functional?

Each requirement placed on a recombinant protein will affect the choice of expression system. If a protein is to be used only to prepare antibody, it need not be soluble or active, and the production of inclusion bodies (aggregates of improperly folded protein) in *E. coli* may be all that is needed. Alternatively, if a protein's biological activity will be assayed, or if it is to be used in structural studies (NMR, crystallography, etc.), a properly folded and soluble form will be required.

Will Structural Changes (Additional or Fewer Amino Acids) Affect Your Application?

Depending on the way that a gene is inserted in an expression vector, additional sequences may be added to the clone, and these may lead to extra amino acid residues at the N- or C-termini of the final expressed protein. In many cases these will have no deleterious effect, but if structural studies or precise comparisons to a native protein are to be done, it is wise to eliminate amino acids added by cloning steps. PCR amplification is the most commonly used method to generate inserts for expression, and proper design of PCR primers can eliminate most or all additional residues in the protein.

Is the Sequence of Your Protein Recognized by Specific Proteases?

If you plan to express your gene in a fusion vector that provides an internal protease cleavage site for removal of the affinity tag (discussed below), check that your native protein is not recognized by the protease. Most proteases are highly specific, but thrombin has a variety of secondary cleavage sites (Chang, 1985).

Advertisements for Commercial Expression Vectors Are Very Promising. What Levels of Expression Should You Expect?

There are several systems available for protein expression in mammalian, insect, yeast, and *E. coli*. While it is impossible to predict the yields of protein from these systems for any given protein, some rough guidelines can be given. For any vector it is possible that no expression will be seen! Reported yields in stably transfected mammalian cells are in the range of 1 to 100 $\mu\text{g}/10^6$

cells. Insect cell systems will yield between 5 and 200mg/L of culture (Schmidt et al., 1998), *Pichia* can produce up to 250mg/L (Eldin et al., 1997), and reported yields in *E. coli* range from 50 μ g to over 100mg/L. Usually yields of from 1 to 10mg/L can be expected from *E. coli*. Higher yields, up to a gram or more per liter, can be had using fermentation vessels where oxygen and pH levels can be controlled throughout the cell growth. The above-mentioned values are guidelines; they are entirely dependent on the protein to be expressed. It is always best to test one or more systems in parallel to select the best solution.

Nonbiological synthesis of protein is now possible as an alternative to production in a host organism (Kochendoerfer and Kent, 1999). Oligopeptides are synthesized and then assembled by chemical ligation to give full-length protein. The method has the potential to synthesize gram quantities of >30kDa proteins, and such preparations would of course be free of host contaminants that might interfere with function or use in diagnostic or therapeutic applications. Unfortunately, chemical synthesis of proteins is not widely available.

Which *E. coli* Strain Will Provide Maximal Expression for Your Clone?

The choice of an expression host depends on the promoter system to be used. Promoters that depend on *E. coli* RNA polymerase can be expressed in most common cloning strains, while T7 promoter vectors must be used in *E. coli* that co-express T7 RNA polymerase (e.g., strains that contain the DE3 lysogen) (Dubendorff and Studier, 1991). Strains that are protease deficient (Bishai, Rappuoli, and Murphy, 1987) or overexpress chaparones have been shown to be useful for some proteins (Georgiou and Valax, 1996; Gilbert, 1994). At a minimum, a recombination deficient strain is advisable. Vendors of the commercially available *E. coli* expression vectors generally will recommend a host for use in expression. As with many questions related to protein expression, the results will depend on the nature of the protein of interest. A given gene may give high yields of intact protein in most strains, while the next would show no product except in a protease deficient host.

Why Should You Select a Fusion System?

Increased Yields

There are several reasons that one would choose to use a fusion system. Translational initiation from the amino terminal fusion

partner may be more efficient than the start contributed by the protein of interest, so larger amounts of protein can be obtained as a fusion. In addition smaller proteins (<20kDa), or sub-fragments of larger ones often benefit from association with a stable fusion partner, due in part to improved folding or protection from proteolysis. Fusion with GST, MBP, and thioredoxin may be useful for this purpose.

Simplified Purification and Detection

Most of the commonly available fusion partners double as affinity tags, and these make isolation of the protein of interest relatively simple. Protein can often be purified to >90% in a single step. In contrast to conventional chromatographic techniques, little or no information about the sequence, pI, or other physical characteristics of the protein is needed in order to perform the purification. Novice chromatographers or those who have not developed methods for purification of the native protein are advised to begin with an affinity system.

Detection of fusion proteins is a simple matter, since antibodies and colorimetric substrates are available for several of the more common fusion partners. Thus, if there is no established method to detect the protein, detection of the fusion partner can be the most convenient way to assay for the presence of the protein in cells and throughout purification and assay of the protein of interest.

When Should You Avoid a Fusion System?

Since affinity tags make purification relatively simple, and tags can be removed by proteolytic cleavage, use of a tag usually makes sense. If, on the other hand, a nonfusion vector has been used in earlier work, and one wishes to compare results with older data, use the nonfusion system. If there is an established method for purification and a biochemical assay or antibody available to detect the protein of interest, an affinity partner or tag for detection may simply be unnecessary. Ask again what use the protein will be put to. If the end application is likely to be sensitive to the presence of the tag (e.g., NMR, crystallography, therapeutics), and other conditions above are met, there is reason to avoid the tag.

If a fusion affinity tag is desired, several are available. Table 15.2 summarizes some of the characteristics of the most widely used fusion partners.

Table 15.2 Commercially Available Fusion Systems

Tag	Tag Size	Purification	Detection	Cleavage
Calmodulin/CBP	(CBP, 4 kDa)	Calmodulin-agarose EGTA for elution	Biotinylated calmodulin and streptavidin alkaline phosphatase	Thrombin, enterokinase
Chitin binding domain (CBD)	<i>Bacillus circulans</i> chitin binding domain (CBD, 52 amino acid residues)	chitin beads	Anti-CBD antibody	Used with intein. On-column cleavage is induced at 4°C by DTT or 2-mercaptoethanol.
E-tag	1.4 kDa	Anti-E sepharose	Anti-E antibodies	NA
FLAG®	1 kDa	Anti-Flag resin	Anti-FLAG antibodies	Enterokinase
Glutathione S-transferase GST	26.5 kDa homodimer GST forms a 58 kDa homodimer with two GSH binding sites. The affinity of the enzyme for GSH is approximately 0.1 μM.	Glutathione sepharose/ Glutathione Agarose	Anti-GST antibodies, CDNB substrate	Thrombin Factor Xa PreScission™ protease
HA (hemagglutinin)	~1 kDa YPYDVPDYA	NA	Anti-HA antibodies	
His ₆	1 kDa	NTA-agarose, Iminodiacetic acid-sepharose	Anti-His ₆ antibodies	Enterokinase, if desired
Maltose binding protein	42.5 kDa K_d of MBP for maltose is 3.5 μM; for maltotriose, 0.16 μM (Miller et al., 1983)	Amylose beads	Anti-MBP	Factor Xa
Myc tag	10 amino acids from human c-Myc EQKLISEEDL	Anti-Myc antibody resin	Anti-Myc antibodies (9E10)	NA
Nus-tag	<i>E. coli</i> NusA protein, 495 amino acids	NA	None	Thrombin
Pinpoint™	12.5 kDa peptide biotinylated <i>in vivo</i> (Samols et al., 1988)	Monomeric avidin resin (SoftLink™ soft release avidin resin)	Avidin/strep tavidin conjugates	Factor Xa
S-tag	15 amino acid peptide (S-tag) with strong affinity ($K_d = 10^{-9}$ M) for a 104 amino acid fragment of	S-protein agarose beads	S-protein FITC conjugate	Thrombin, enterokinase

Table 15.2 (Continued)

Tag	Tag Size	Purification	Detection	Cleavage
Strep-tag	pancreatic ribonuclease A. A 10 amino acid sequence that binds streptavidin	Streptavidin bead	Streptavidin conjugates	
Z-domain	Two Z domains add a 14 kDa peptide	IgG-sepharose		Factor Xa

Susceptibility To Cleavage Enzymes

As discussed below, some fusion systems allow for the removal of the affinity tag by specific proteolytic or chemical cleavage. Before beginning any experiment, examine the sequence of the protein to be cloned and expressed. The protein of interest may have a binding site for one of the proteases listed in Table 15.3, and if so, this site should be avoided, or a different expression system might be required. Most proteases used for cleavage of fusion protein are quite specific, with theoretical frequencies of 10^{-6} . However, it is best to check as a matter of course.

Is It Necessary to Cleave the Tag off the Fusion Protein?

For many proteins, cleavage is not needed. If the goal of the work is to raise an antibody, the whole fusion protein can be used successfully as antigen—provided that antibodies to the tag do not interfere in the application. If, on the other hand, the protein is to be used in structural studies, or where the function of recombinant protein will be compared with native protein, it may be necessary to remove the fusion tag.

Systems have been developed that use chemical (Nilsson et al., 1985) or specific proteolytic cleavage to separate the protein of interest from the fusion tag. The proteases have the advantage that cleavage is done at near neutral pH and at 4 to 37°C. In addition to proteolytic cleavage, the use of self-splicing inteins has been developed and commercialized by New England Biolabs. In this latter case fusion proteins with chitin-binding domain are bound to high molecular weight chitin chromatography media and incubated in the presence of a reducing agent, generally overnight. Protein splicing takes place, leaving the protein of interest in the flow through, while chitin and the spliced peptide remain bound.

Table 15.3 Characteristics of Popular Fusion Protein Cleavage Enzymes

Protease	Cleavage Site	Comment
Thrombin	?VPR^GS secondary cleavage sites exist; (Chang, 1985)	Widely used, works at 1:1000–1:2000 mass ratio relative to target protein. Purified from bovine sources and may include other proteins.
Factor Xa protease	IEGR^	Leaves defined N-terminus. Works at 1:500–1:1000 mass ratio relative to target protein. Recognition site with proline immediately following Arg residue will not be cleaved.
Enterokinase	DDDDK^	Leaves defined N-terminus. Recombinant.
rTEV	ENLYFQ^G	Recombinant endopeptidase from the Tobacco Etch Virus.
Intein-mediated self-cleavage		No added protease required. Leaves defined N-terminus
PreScission protease	LEVLFQ^GP	Rhinoviral 3C protease expressed as GST fusion protein. Optimal activity at 4°C.

Recognition sites for enzymes commonly used to cleave fusion proteins, and their advantages/disadvantages are listed in Table 15.3

Will Extra Amino Acid Residues Affect Your Protein of Interest after Digestion?

Depending on the protease, and the way in which the protein of interest was cloned in the expression vector, there may be one or more nonnative residues left at the amino terminal of the protein of interest following cleavage. Whether or not this poses a problem depends entirely on the protein and the use to which it will be put. Even the most demanding applications may not be negatively affected by the presence of extra amino terminal residues. Wherever possible, it is best to design a cloning strategy that at least minimizes the number of these residues, and if relatively innocuous residues (e.g., glycine, serine) can be introduced, all the better.

WORKING WITH EXPRESSION SYSTEMS

What Are the Options for Cloning a Gene for Expression?

In some cases the protein of interest is already cloned in another vector, for example, in a clone isolated from a cDNA

expression library. If the frame of the insertion is known, and compatible restriction sites are found in the expression vector(s) selected, the insert can be cloned directly. In some cases excision from a lambda vector can generate a plasmid vector ready for expression of the insert, without any manipulation at all.

More commonly PCR is used to amplify the target sequence using oligonucleotide primers that have 15 to 20 bases of homology with the 5' and 3' ends of the target. These primers will have in addition tails that encode restriction enzyme sites compatible with the expression vector. The PCR products can be digested with the appropriate restriction enzymes, purified, and ligated into an appropriately prepared vector.

The efficiency of cloning can be improved if two different restriction enzyme sites are available. This will allow for directional cloning of inserts into the vector, and all of the clones screened should have the insert in the desired orientation. Please refer to Chapter 9, "Restriction Endonucleases" for a discussion on double digestion strategies. If PCR is used to generate the insert, then primers must be designed appropriately. It is important to leave 4 to 6 random bases at the 5' end of each PCR primer. These provide a spacer at the ends of the PCR product and allow the restriction enzymes to digest the DNA more efficiently. While *in vitro* ligation is still the most widely used method, ligation independent cloning (LIC) (Li and Evans, 1997) has the advantage that no DNA ligase is required (though an exonuclease activity is), and efficiencies are comparable to those obtained with conventional ligation with T4 DNA ligase.

Is Screening Necessary Prior to Expression?

There are no guarantees that the gene to be expressed will be present in the cell after transformation. As discussed above, most expression vectors are prone to produce small amounts of the protein even in the absence of inducing agent, which can prove toxic to the host. Alternatively, host cells can cause deletions and rearrangements in the expression vector. Either way, it is usually a very good idea to confirm the presence of the inserted gene prior to expression experiments.

Unless a library of clones is to be prepared, the efficiency of ligation and transformation is rarely an issue. Screening of a dozen clones for the presence of an insert should be sufficient to identify one or more positive candidate clones.

The first step is generally to prepare several plasmid DNA minipreps and digest the DNA with the same enzyme(s) used in

cloning to generate the insert. Products should be analyzed by agarose gel electrophoresis to determine if DNA of the predicted size was inserted in the vector. As an alternative, PCR can be done using as template a small scraping from a colony on the plate. Amplification of the plasmid DNA contained in the cells using the same primers used in cloning, or primers that anneal to flanking vector sequences, should show a band of the predicted size. This latter method does not confirm the presence of the restriction sites used in cloning, but has the advantage of being rapid.

Once the presence of an insert of the correct size is confirmed, the DNA sequence at the cloning junctions should be determined. It is not uncommon for a primer sequence to be synthesized with an error—whether by faulty design or at the hands of the oligo supplier. DNA sequencing to confirm the cloning junctions should be done in parallel with a small-scale expression experiment, in which a 1 to 2 ml culture is grown and induced according to a standard protocol. It is important to include a culture that is transformed with the parent expression vector as a negative control in this screening experiment. Following centrifugation, the cell pellet should be suspended in SDS-PAGE loading buffer, and a small amount loaded on an SDS-acrylamide gel. The viscosity of the whole cell lysate (caused by the release of genomic DNA) may make gel loading difficult. However, addition of extra 1× loading buffer, DNaseI (10 μg/ml), extended heating of the sample, or sonication should alleviate the problem.

After electrophoresis, the gel should be stained (e.g., Coomassie Blue) to visualize the proteins in the whole cell lysate. If expression is good, an induced band will clearly be seen at the predicted molecular weight, and this will be absent in the no-insert control culture. If no band is visible and the restriction digestion/DNA sequencing data indicate that all is well, don't despair. Perform an immunoblot of an SDS-acrylamide gel. Screen for the presence of the protein of interest or use an antibody directed against the affinity or epitope tag if one has been used. Use of both N- and C-terminal specific antibodies is ideal in troubleshooting. Be sure to include both positive and negative controls in the immunoblot. Alternatives to immunoblotting include ELISA or specific biochemical assays for the protein of interest.

If an antibody is not available for Western blotting, and you have a procedure to purify your protein, attempt the purification. This can visualize a protein that is present in quantities insufficient to stand out on a PAGE gel of a total cell extract.

Once expression of a protein of the predicted molecular weight is found, minimize propagation of the cells. Serial growths under

conditions that permit expression may lead to plasmid loss or rearrangements.

Once analysis is complete, glycerol stocks of positive clones should be prepared. This can be done by streaking culture residue from the DNA miniprep on a plate to get a fresh colony, by reusing the colony that was originally picked, or by re-transforming *E. Coli* with isolated miniprep DNA. In either case a fresh colony should be used to prepare a 2 to 4ml log-phase culture for the purpose of making a glycerol stock. Be sure to keep protein expression repressed during this step by reduced temperature, use of minimal medium, or adapting it to the vector in use.

What Aspects of Growth and Induction Are Critical to Success?

Aeration, Temperature

The best expression results are had when cultures are grown with sufficient aeration and positive selection for the plasmid. For small-scale experiments, use 2ml of medium (e.g., LB, SOC or 2XTY) in a 15 ml culture tube. Vigorous shaking (>250rpm) should be used to maintain aeration. Appropriate antibiotics, such as ampicillin should be added to recommended levels. At larger scales, Ehrlenmeyer flasks should be used. Flasks with baffles improve aeration and $\frac{1}{8}$ to $\frac{1}{2}$ of the total volume of the flask should be occupied by medium. Good results may be obtained using 250ml to 1 L in a 2L baffled flask.

Scaling Up

When scaling up growth, monitor the light scattering at 590 or 600nm. Note that a culture with OD₆₀₀ of one corresponds to about 5×10^8 cells/ml, though this number will vary depending on the strain of *E. coli* used. Two rules of thumb are particularly important: minimize the time in each stage of growth, and monitor both cell density and protein expression at each stage.

From a colony or glycerol stock, begin a small overnight culture (e.g., 2–5 ml) in a selective medium under conditions that repress expression. Don't allow the culture to overgrow. This starter culture is then used to inoculate a larger volume of medium at a volume ratio of about 1:100 (pre-warming the media is a good idea). Monitor the growth by absorbance at 600nm, and keep the cell density low (OD₆₀₀ below 1). Once the growth has been scaled to give sufficient starter for the final growth vessel, make an inoculum of about 1%. Monitor the OD every 30 minutes or so, and remove aliquots for analysis by SDS-PAGE, immunoblotting, or

functional assay. After an initial lag following the inoculation, the density of the cells should double every 20 to 40 minutes. A graph of the OD coupled with an immunoblot is very useful in selecting optimal conditions for the growth. Once the culture reaches a late log phase (usually about OD₆₀₀ of 0.8–1.2), induction is done by the addition of the appropriate inducing agent. Continue to monitor growth and take aliquots. It is not unusual that cells expressing a foreign protein will either stop growing or show a 10% to 20% decrease in density following induction. While it is common to grow for 1 to 3 hours postinduction prior to harvest, this induction period can vary depending on temperature and other conditions. So it is best determined empirically.

What Are the Options for Lysing Cells?

E. coli are easily broken by several methods including decavitation, shearing, and the action of freeze–thaw cycles. The choice of method depends on the scale of growth, and the type of equipment available (reviewed in Johnson, 1998). For most lab-scale experiments, sonication, or freeze–thaw will be the most practical choices. Ultrasonic disruptors are available from many vendors, but all operate on the conversion of electrical energy through piezoelectric transducers into ultrasonic waves of 18 to 50kHz. The vibration is transferred to the sample by a titanium tip, and the energy released causes decavitation and shearing of the cells. Several models are available that are microprocessor controlled, programmable, and allow very reproducible cell lysis. It is important to keep the sample on ice and avoid frothing. This latter problem is caused by a probe that is not immersed sufficiently in the sample, or by excessive power. If bubbles begin forming and accumulating on the surface, stop immediately, reposition the probe, and reduce output. Once a sample has been turned to foam, sonication will be ineffective, and there is little to do but start again. Even if frothing is not seen, treatment beyond that needed to cause cell lysis can result in physical damage to the protein of interest. The addition of protease inhibitors to the cell suspension immediately prior to cell lysis is an important precaution, and several commercial cocktails are available for this purpose.

Freeze–thaw, particularly in conjunction with lysozyme and DNase treatment, is one of the mildest procedures to break *E. coli*. Cells are simply resuspended in buffer (PBS, Tris-pH 8.0) containing 10μg/ml hen egg lysozyme, and the sample is cycled between a dry ice–alcohol bath and a container of tepid water. Generally, 5 to 10 cycles is sufficient to break nearly all of the cells.

As the cells lyse and DNA is liberated, it may be necessary to add DNase to 10 $\mu\text{g/ml}$ to reduce the viscosity of the preparation. Commercial or homemade detergent preparations including *N*-octylamine are also very effective at lysing cells and simple to use.

Whatever method is used, lysis should be monitored. Microscopic examination is the best option. Retain some of the starting suspension, and compare to the lysate. Phase contrast optics will permit direct visualization, though staining can be used as well. Lysis will be evidenced by a slight darkening of the suspension, or clearing, and under the microscope, cells will be broken with membrane fragments or small vesicles present.

Other physical lysis methods include the use of a French Press, Manton-Gaulin, and other devices that place cells under rapid changes of pressure or shear force. These are very effective and reproducible, but generally, they are best used when the original culture volume is $>1\text{ L}$, since most of these cell disruptors have minimum volume requirements.

TROUBLESHOOTING

No Expression of the Protein

If one has checked for small-scale expression as discussed above, there should have been a detected band on a stained gel or immunoblot. If neither are seen, sequencing of the cloning junctions and entire insert should be undertaken to confirm that no frame shifts, stop codons, or rearrangements have occurred. Purification can be tried in parallel to see if even very low levels are made. A slight band on SDS-PAGE of the expected protein will make clear that the cloning went as planned, but the biology of expression is at fault. Varying temperature, time of induction, and the type of plasmid or fusion system can all be tried. In the end some proteins may not express well in *E. coli*, and they should be tried in other organisms.

The Protein Is Expressed, but It Is Not the Expected Size Based on Electrophoretic Analysis

On SDS-PAGE the net charge on the protein of interest will affect mobility. Highly charged proteins will tend to bind less SDS and will have retarded mobility. Proteins rich in proline may also exhibit dramatically slower mobility in SDS-PAGE. If the protein has a calculated pI in the range of 5 to 9, and is not strongly biased in amino acid composition, then a protein that shows multiple

bands or a strong species far from the predicted molecular weight is likely due to something other than an artifact of SDS-PAGE. Probing immunoblots with the appropriate antibodies to N- and/or C-terminal tags of the protein is particularly useful at this stage. Try to identify the halves or pieces of the protein on stained gels and immunoblots to locate likely points in the coding sequence where proteolytic cleavage and/or translation termination may occur. Cleavage at the junction between the protein of interest and the fusion partner (if any) that is used is often seen. Addition of protease inhibitors should be routine in all work, and protease-deficient strains should be tried in parallel or as a next step. If these measures fail, try re-cloning in another vector with a different fusion tag or tags, and promoter.

The Protein Is Insoluble. Now What?

Many heterologous proteins expressed in *E. coli* will be found as insoluble aggregates that are failed folding intermediates (Schein, 1989). Such inclusion bodies are seen as opaque areas in micrographs of *E. coli* that express the protein of interest. Depending on the purpose of expression, the production of inclusion bodies may be a welcome occurrence. If for example, the recombinant protein is to be used solely for the production of antibodies, inclusion bodies may be isolated to high purity by differential centrifugation and used directly as an antigen. If the protein is relatively small, the inclusion bodies may be isolated as above, and refolded with good efficiency. Other (particularly large) proteins will not refold well, and if production of functional protein is required, then an alternative must be found. Before proceeding, it is best to answer the following questions.

Are You Sure Your Protein Is Insoluble?

A first consideration is whether the protein is truly insoluble, or the cells were simply not lysed. Here is where microscopic examination will be of great use. Examine a cell lysate under phase contrast microscopy or after staining. Are intact cells visible? After it sediments, is the pellet large and similar in appearance to the original cell pellet? Is the post-lysis supernatant clear? Any of the above may indicate that cells are not completely disrupted. The protein of interest may be soluble but trapped in intact cells.

If cells are lysed as measured by microscopy, analyze whole cell lysate, clarified lysate, and post-lysis pellet by SDS-PAGE, followed by staining or immunoblotting. If cells are lysed as mea-

sured by microscopy, and the protein of interest is found in the post-lysis pellet, it is likely that it is being made in an insoluble form. While most use a relatively low-speed centrifugation step at around $10,000 \times g$, it is best to do a $100,000 \times g$ spin to sediment all aggregates before drawing any conclusion about insolubility. Another indication is microscopic examination of cells under high power ($>400\times$). If inclusion bodies are being made, and expression levels are high, optically dense areas in the *E. coli* cells will be seen. These inclusion bodies may occupy more than half of the cell.

Must Your Protein Be Soluble?

The accumulation of proteins in inclusion bodies is not necessarily undesirable. Insolubility has three important advantages:

1. Inclusion bodies can represent the highest yielding fraction of target protein.
2. Inclusion bodies are easy to isolate as an efficient first step in a purification scheme. Nuclease-treated, washed inclusion bodies are usually 75% to 95% pure target protein.
3. Target proteins in inclusion bodies are generally protected from proteolytic breakdown.

Isolated inclusion bodies can be solubilized by a variety of methods in preparation for further purification and refolding. If the application is to prepare antibodies, inclusion bodies can be used directly for injection after suspension in PBS and emulsification with a suitable adjuvant (e.g., Fischer et al., 1992). If the target protein contains a his₆-tag, purification can be performed under denaturing conditions. The purified protein can be eluted from the resin under denaturing conditions and then refolded.

Solubility Is Essential. What Are Your Options?

Prevent Formation of Insoluble Bodies

A number of approaches have been used to obtain greater solubility, including induction of protein expression at 15 to 30°C (Burton et al., 1991), use of lower concentrations of IPTG (e.g., 0.01–0.1 mM) for longer induction periods, and/or using a minimal defined culture medium (Blackwell and Horgan, 1991).

Solubilize and Refold

Solubilization and refolding methods usually involve the use of chaotropic agents, co-solvents or detergents (Marston and

Hartley, 1990; Frankel, Sohn, and Leinwand, 1991; Zardeneta and Horowitz, 1994). A strategy that has been successful for some proteins is to express as a his₆-tagged fusion, bind under denaturing conditions, and refold while protein is still bound to the resin by running a gradient from 6M to 0M guanidine-HCl in the presence of reduced (GSH) and oxidized (GSSH) glutathione. Once folding has occurred, elution is done with imidazole as usual. Some researchers enhance refolding of enzymes by the addition of substrate or a substrate analogue during gradual removal of denaturant by dialysis (Zhi et al., 1992; Taylor et al., 1992).

The Protein Is Made, but Very Little Is Full-Length; Most of It Is Cleaved to Smaller Fragments

It is important to distinguish among proteolytic breakdown, translation termination, and cryptic translation start sites within the gene of interest. Proteolytic breakdown is most likely to occur at exposed domains of the protein. Examine the pattern of breakdown products by SDS-PAGE, estimate their sizes, and compare the result with the predicted amino acid sequence. Keep an eye out for bends or surface-exposed regions, and any sequences that conform to those for known proteases. While protease inhibitors such as PMSF should be present in the sample prior to cell lysis, expand the group of protease inhibitors and test their effect. Also consider the pattern of expression seen when growth is monitored before and after induction. If there is a switch between intact and fragmented protein after induction, it is likely that proteolysis is the culprit.

Translation Termination

There is little clear-cut evidence for inappropriate translation termination, but in at least one case a stretch of 20 serine residues was suggested to cause premature termination in *E. coli* (Bula and Wilcox, 1996). If a truncated protein is definitely seen, DNA sequencing in the expected termination region should be done to confirm that no cryptic stop codons exist.

Cryptic translation initiation may be seen as well (Preibisch et al., 1988). Cryptic translation initiation can occur within an RNA coding sequence when a sequence resembling the ribosome binding site (AAGGAGG) occurs with the appropriate spacing (typically 5 to 20 nucleotides upstream of an AUG (Met) codon). These smaller products can be problematic when attempting to purify full-length proteins. If some expression of full-length

protein is seen, a useful strategy may be to try dual tag affinity purification, in which the gene of interest is expressed in a vector that encodes two affinity tags, one each at the C- and N-termini. Sequential purification using both affinity tags can give reasonable yields of full-length protein whatever the original cause (Kim and Raines, 1994; Pryor and Leiting, 1997).

Your Fusion Protein Won't Bind to Its Affinity Resin

A lysate is produced, and contacted with the affinity medium. The protein of interest is present in the cell and clarified lysate, as shown by SDS-PAGE, but after purification of the lysate over the medium, all of the protein is found in the flow-through. The presence of a large amount of protein in the eluate after an attempt to bind to the affinity medium does not prove an inability to bind. If there is a very large excess of protein, it may appear that none is binding, when in fact the column has simply been overloaded. Try to wash and elute the protein from the affinity medium before drawing a conclusion. One simple test is to remove 10 to 50 μ l of the purification medium after binding and washing, and then boil the sample in an equal volume of 1 $\times$ SDS-PAGE loading dye. Gel analysis may show binding of the protein to the resin. Consideration of the amount loaded on the column and the expected capacity of the purification medium will sort out the various causes. If in fact expression is clearly seen in the lysate applied to the purification medium, there are other explanations:

1. The affinity medium was not equilibrated properly, or the protein folded to mask the residues responsible for binding to the affinity medium. Purification in the presence of detergents (e.g., 0–1% Tween-20), or mild chaotropes (e.g., 1–3M guanidine-HCl or urea) may unmask these residues and enable binding.
2. Your fusion protein won't elute from its affinity resin. Protein may apparently bind to the resin, as measured by the presence of an SDS-PAGE gel band after boiling a sample of the washed resin. Little or no protein of interest may be eluted, however, when the loaded resin is contacted with eluting agent. In this latter case the protein may interact nonspecifically with the base matrix, or the protein precipitated during contact with the resin and is trapped. Addition of detergent, of varying ionic strength and pH, may improve the situation.

Your Fusion Protein Won't Digest

If expression is otherwise good, and the protein is not digested to *any* extent, one should confirm by DNA sequencing that the protease site is intact. Checking the activity of the protease in parallel experiments using a known and well-behaved protein will give some confidence that the protease itself is not to blame. If the site is present, the protease has activity, and buffer conditions are close to those specified for the protease, it may be that the fusion protein folds so that the protease site is inaccessible. Additives that alter the structure slightly, including salts and detergents may unmask the site; see Ellinger et al. (1991). Alternatively, recloning to create a flexible linker flanking the protease site has been shown to increase the efficiency of digestion with Thrombin (Guan and Dixon, 1991) and presumably other proteases.

Cleavage of the Fusion Protein with a Protease Produced Several Extra Bands

Cryptic Sites

The specificity of any protease is inferred from its natural substrates, and there is reason to believe that cryptic sites are also cleaved. (Nagai, Perutz, and Poyart, 1985; Eaton, Rodriguez, and Vehar, 1986; Quinlan, Moir, and Stewart, 1989; Wearne, 1990).

Excess Protease

If multiple bands are seen by SDS-PAGE, a titration of the amount, time and temperature of digestion should be done. Often reducing time or temperature will minimize cleavage at secondary sites, while retaining digestion at the desired site.

Extra Protein Bands Are Observed after Affinity Purification

E. coli host chaperone protein GroEL, with an apparent molecular weight of about 57 to 60kDa on SDS-PAGE, is often found to co-purify with a protein of interest (Keresztessy et al., 1996). This may be caused by misfolding or by a recombinant protein that is trapped at an intermediate folding stage. High salt concentration (1–2 M), non-ionic detergents, and ligand or co-factors (e.g., ATP or GTP) may be effective in removing chaperones from the protein of interest. Often chaperones and other contaminating proteins are seen following affinity purification; they are best removed by conventional chromatography such as ion exchange.

Their co-purification can be minimized by inducing the culture at a lower density (e.g., $OD_{600} = 0.3$ vs. 1.0) or by reducing temperature.

Must the Protease Be Removed after Digestion of the Fusion Protein?

The removal of the protease is not necessary for many applications. Generally, protease is added at a ratio of 1:500 or lower relative to the protein of interest, so protease may not interfere with downstream applications. Biochemical assays and antibody production may not require removal, while structural studies, or assays where other proteins are added to the protein of interest in a biochemical assay indicate that a further purification be performed.

The commonly used serine proteases, thrombin and factor Xa, can be removed from a reaction mixture by contacting the digested protein/protease with an immobilized inhibitor such as benzamidine-sepharose (Sundaram and Brandsma, 1996). This purification is not complete due to the equilibrium binding of the inhibitor to the protease, but the majority of the protease can be removed in this way. Better yet, a different purification method like ion-exchange or hydrophobic interaction chromatography can be used to separate the protein of interest from both the protease and other cleavage products including the affinity tag.

Some commercially available proteases (Table 15.3) include affinity tags that can be used effectively to remove the protease from the sample. Biotinylated thrombin can be removed with high efficiency due to the extreme affinity of biotin for avidin or streptavidin-agarose beads. Other proteases containing affinity tags include PreScission protease; a fusion of GST with human rhinoviral 3C protease.

BIBLIOGRAPHY

- Beck von Bodman, S., Schuler, M. A., Jollie, D. R., and Sligar, S. G. 1986. Synthesis, bacterial expression, and mutagenesis of the gene coding for mammalian cytochrome b5. *Proc. Nat. Acad. Sci. U.S.A.* 83:9443–9447.
- Bessette, P. H., Aslund, F., Beckwith, J., and Georgiou, G. 1999. Efficient folding of proteins with multiple disulfide bonds in the *Escherichia coli* cytoplasm. *Proc. Nat. Acad. Sci. U.S.A.* 96:13703–13708.
- Bishai, W. R., Rappuoli, R., and Murphy, J. R. 1987. High-level expression of a proteolytically sensitive diphtheria toxin fragment in *Escherichia coli*. *J. Bact.* 169:5140–5151.
- Blackwell, J. R., and Horgan, R. 1991. A novel strategy for production of a highly expressed recombinant protein in an active form. *FEBS Lett.* 295:10–12.

- Boyd, D., and Beckwith, J. 1990. The role of charged amino acids in the localization of secreted and membrane proteins. *Cell* 62:1031–1033.
- Buell, G., Schulz, M. F., Selzer, G., Chollet, A., Movva, N. R., Semon, D., Escanez, S., and Kawashima, E. 1985. Optimizing the expression in *E. coli* of a synthetic gene encoding somatomedin-C (ICF-I). *Nucl. Acids Res.* 13:1923–1938.
- Bula, C., and Wilcox, K. W. 1996. Negative effect of sequential serine codons on expression of foreign genes in *Escherichia coli*. *Prot. Expr. Purif.* 7:92–103.
- Burton, N., Cavallini, B., Kanno, M., Moncollin, V., and Egly, J. M. 1991. Expression in *Escherichia coli*: Purification and properties of the yeast general transcription factor TFIID. *Prot. Expr. Purif.* 2:432–441.
- Chang, J.-Y. 1985. Thrombin specificity: Requirement for apolar amino acids adjacent to the cleavage site of polypeptide substrate. *Eur. J. Biochem.* 151:217–224.
- Chen, G. F., and Inouye, M. 1990. Suppression of the negative effect of minor arginine codons on gene expression; preferential usage of minor codons within the first 25 codons of the *Escherichia coli* genes. *Nucl. Acids Res.* 18:1465–1473.
- Derman, A. I., Prinz, W. A., Belin, D., and Beckwith, J. 1993. Mutations that allow disulfide bond formation in the cytoplasm of *Escherichia coli*. *Science* 262:1744–1747.
- Dodd, I., Mossakowska, D. E., Camilleri, P., Haran, M., Hensley, P., Lawlor, E. J., McBay, D. L., Pindar, W., and Smith, R. A. 1995. Overexpression in *Escherichia coli*: Folding, purification, and characterization of the first three short consensus repeat modules of human complement receptor type 1. *Prot. Expr. Purif.* 6:727–736.
- Dubendorff, J. W., and Studier, F. W. 1991. Controlling basal expression in an inducible T7 expression system by blocking the target T7 promoter with lac repressor. *J. Mol. Biol.* 219:45–59.
- Eaton, D., Rodriguez, H., and Vehar, G. A. 1986. Proteolytic processing of human factor VIII: Correlation of specific cleavages by thrombin, factor Xa, and activated protein C with activation and inactivation of factor VIII coagulant activity. *Biochem.* 25:505–512.
- Eldin, P., Pauza, M. E., Hieda, Y., Lin, G., Murtaugh, M. P., Pentel, P. R., and Pennell, C. A. 1997. High-level secretion of two antibody single chain Fv fragments by *Pichia pastoris*. *J. Immunolog. Meth.* 201:67–75.
- Ellinger, S., Mach, M., Korn, K., and Jahn, G. 1991. Cleavage and purification of prokaryotically expressed HIV gag and env fusion proteins for detection of HIV antibodies in the ELISA. *Viol.* 180:811–813.
- Etchegaray, J. P., and Inouye, M. 1999. Translational enhancement by an element downstream of the initiation codon in *Escherichia coli*. *J. Biol. Chem.* 274:10079–10085.
- Fischer, L., Gerard, M., Chalut, C., Lutz, Y., Humbert, S., Kanno, M., Chambon, P., and Egly, J. M. 1992. Cloning of the 62-kilodalton component of basic transcription factor BTF2. *Science* 257:1392–1395.
- Frankel, S., Sohn, R., and Leinwand, L. 1991. The use of sarkosyl in generating soluble protein after bacterial expression. *Proc. Nat. Acad. Sci. U.S.A.* 88:1192–1196.
- Georgiou, G., and Valax, P. 1996. Expression of correctly folded proteins in *Escherichia coli*. *Curr. Opin. Biotechnol.* 7:190–197.
- Gilbert, H. F. 1994. Protein chaperones and protein folding. *Curr. Opin. Biotechnol.* 5:534–539.
- Gold, L., Pribnow, D., Schneider, T., Shinedling, S., Singer, B. S., and Stormo, G. 1981. Translational initiation in prokaryotes. *Ann. Rev. Microbiol.* 35:365–403.
- Guan, K. L., and Dixon, J. E., 1991. Eukaryotic proteins expressed in *Escherichia coli*: An improved thrombin cleavage and purification procedure of fusion proteins with glutathione S-transferase. *Anal. Biochem.* 192:262–267.

- Hirel, P. H., Schmitter, M. J., Dessen, P., Fayat, G., and Blanquet, S. 1989. Extent of N-terminal methionine excision from *Escherichia coli* proteins is governed by the side-chain length of the penultimate amino acid. *Proc. Nat. Acad. Sci. U.S.A.* 86:8247–8251.
- Hostomsky, Z., Smrt, J., and Paces, V. 1985. Cloning and expression in *Escherichia coli* of the synthetic proenkephalin analogue gene. *Gene* 39:269–274.
- Hua, Z., Wang, H., Chen, D., Chen, Y., and Zhu, D. 1994. Enhancement of expression of human granulocyte-macrophage colony stimulating factor by argU gene product in *Escherichia coli*. *Biochem. Mol. Biol. Int.* 32:537–543.
- Izard, J. W., and Kendall, D. A. 1994. Signal peptides: Exquisitely designed transport promoters. *Mol. Microbiol.* 13:765–773.
- Johnson, B. 1998. Breaking up isn't hard to do: A cacophony of sonicators, cell bombs, and grinders. *Scientist* 12:23.
- Kapust, R. B., and Waugh, D. S. 1999. *Escherichia coli* maltose-binding protein is uncommonly effective at promoting the solubility of polypeptides to which it is fused. *Prot. Sci.* 8:1668–1674.
- Kelley, R. F., and Winkler, M. E. 1990. Folding of eukaryotic proteins produced in *Escherichia coli*. *Genet. Eng. (NY)* 12:1–19.
- Keresztessy, Z., Hughes, J., Kiss, L., and Hughes, M. A. 1996. Co-purification from *Escherichia coli* of a plant beta-glucosidase-glutathione S-transferase fusion protein and the bacterial chaperonin GroEL. *Biochem. J.* 314:41–47.
- Kim, J. S., and Raines, R. T. 1994. Peptide tags for a dual affinity fusion system. *Anal. Biochem.* 219:165–166.
- Kochendoerfer, G. G., and Kent, S. B. 1999. Chemical protein synthesis. *Curr. Opin. Chem. Biol.* 3:665–671.
- Koshiba, T., Hayashi, T., Miwako, I., Kumagai, I., Ikura, T., Kawano, K., Nitta, K., and Kuwajima, K. 1999. Expression of a synthetic gene encoding canine milk lysozyme in *Escherichia coli* and characterization of the expressed protein. *Prot. Eng.* 12:429–435.
- Lathrop, B. K., Burack, W. R., Biltonen, R. L., and Rule, G. S. 1992. Expression of a group II phospholipase A2 from the venom of *Agkistrodon piscivorus piscivorus* in *Escherichia coli*: Recovery and renaturation from bacterial inclusion bodies. *Prot. Expr. Purif.* 3:512–517.
- Lee, N., Zhang, S. Q., Cozzitorto, J., Yang, J. S., and Testa, D. 1987. Modification of mRNA secondary structure and alteration of the expression interferon alpha 1 in *Escherichia coli*. *Gene* 58:77–86.
- Li, C., and Evans, R. M. 1997. Ligation independent cloning irrespective of restriction site compatibility. *Nucl. Acids Res.* 25:4165–4166.
- Looman, A. C., Bodlaender, J., de Gruyter, M., Vogelaar, A., and van Knippenberg, P. H. 1986. Secondary structure as primary determinant of the efficiency of ribosomal binding sites in *Escherichia coli*. *Nucl. Acids Res.* 14:5481–5496.
- Marston, F. A., and Hartley, D. L. 1990. Solubilization of protein aggregates. *Meth. Enzymol.* 182:264–282.
- Miller III, D. M., Olson, J. S., Pflugrath, J. W., and Quioco, F. A. 1983. Rates of ligand binding to periplasmic proteins involved in bacterial transport and chemotaxis. *J. Biol. Chem.* 258:13665–13672.
- Nagai, K., and Thogersen, H. C. 1984. Generation of beta-globin by sequence-specific proteolysis of a hybrid protein produced in *Escherichia coli*. *Nature* 309:810–812.
- Nagai, K., Perutz, M. F., and Poyart, C. 1985. Oxygen binding properties of human mutant hemoglobins synthesized in *Escherichia coli*. *Proc. Nat. Acad. Sci. U.S.A.* 82:7252–7255.
- Nilsson, B., Holmgren, E., Josephson, S., Gatenbeck, S., Philipson, L., and Uhlen,

- M. 1985. Efficient secretion and purification of human insulin-like growth factor I with a gene fusion vector in *Staphylococci*. *Nucl. Acids Res.* 13: 1151–1162.
- Pines, O., and Inouye, M. 1999. Expression and secretion of proteins in *E. coli*. *Mol. Biotechnol.* 12:25–34.
- Preibisch, G., Ishihara, H., Tripier D., and Leineweber, M. 1988. Unexpected translation initiation within the coding region of eukaryotic genes expressed in *Escherichia coli*. *Gene* 72:179–186.
- Pryor, K. D., and Leitig, B. 1997. High-level expression of soluble protein in *Escherichia coli* using a His₆-tag and maltose-binding-protein double-affinity fusion system. *Prot. Expr. Purif.* 10:309–319.
- Quinlan, R. A., Moir, R. D., and Stewart, M. 1989. Expression in *Escherichia coli* of fragments of glial fibrillary acidic protein: Characterization, assembly properties and paracrystal formation. *J. Cell Sci.* 93:71–83.
- Robinson, M., Lilley, R., Little, S., Emtage, J. S., Yarranton, G., Stephens, P., Millican, H., Eaton, M., Humphries, G. 1984. Codon usage can affect efficiency of translation of genes in *Escherichia coli*. *Nucl. Acids Res.* 12:6663–6671.
- Sambrook, J., Fritsch, E. F., and Maniatis, T. 1989. *Molecular Cloning: A Laboratory Manual*, 2nd ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
- Samols, D., Thornton, C. G., Murtif, V. L., Kumar, G. K., Haase, F. C., and Wood, H. G. 1988. Evolutionary conservation among biotin enzymes. *J. Biol. Chem.* 263:6461–6464.
- Schein C. H. 1989. Production of soluble recombinant proteins in bacteria. *Biotechnol.* 7:1141–1149.
- Schmidt, M., Tuominen, N., Johansson, T., Weiss, S. A., Keinänen, K., and Oker-Blom, C. 1998. Baculovirus-mediated large-scale expression and purification of a polyhistidine-tagged rubella virus capsid protein. *Prot. Expr. Purif.* 12:323–330.
- Sorensen, M. A., Kurland, C. G., and Pedersen, S. 1989. Codon usage determines translation rate in *Escherichia coli*. *J. Mol. Biol.* 207:365–377.
- Stader, J. A., and Silhavy, T. J. 1990. Engineering *Escherichia coli* to secrete heterologous gene products. *Meth. Enzymol.* 185:166–187.
- Sundaram, P., and Brandsma, J. L. 1996. Rapid, efficient, large-scale purification of unfused, nondenatured E7 protein of cottontail rabbit papillomavirus. *J. Virol. Meth.* 57:61–70.
- Taylor, M. A., Pratt, K. A., Revell, D. F., Baker, K. C., Sumner, I. G., and Goodenough, P. W. 1992. Active papain renatured and processed from insoluble recombinant propapain expressed in *Escherichia coli*. *Prot. Eng.* 5:455–459.
- Tessier, L. H., Sondermeyer, P., Faure, T., Dreyer, D., Benavente, A., Villeval, D., Courtney, M., and Lecocq, J. P. 1984. The influence of mRNA primary and secondary structure on human IFN-gamma gene expression in *E. coli*. *Nucl. Acids Res.* 12:7663–7675.
- Tobias, J. W., Shrader, T. E., Rocap, G., and Varshavsky, A. 1991. The N-end rule in bacteria. *Science* 254:1374–1377.
- Wearne, S. J. 1990. Factor X_a cleavage of fusion proteins: Elimination of non-specific cleavage by reversible acylation. *FEBS Lett.* 263:23–26.
- Wickner, W., Driessen, A. J., and Hartl, F. U. 1991. The enzymology of protein translocation across the *Escherichia coli* plasma membrane. *Ann. Rev. Biochem.* 60:101–124.
- Wilkinson, D. L., and Harrison, R. G. 1991. Predicting the solubility of recombinant proteins in *Escherichia coli*. *Biotechnol.* 9:443–448.
- Yamane, K., and Mizushima, S. 1988. Introduction of basic amino acid residues after the signal peptide inhibits protein translocation across the cytoplasmic

- membrane of *Escherichia coli*: Relation to the orientation of membrane proteins. *J. Biol. Chem.* 263:19690–19696.
- Yasukawa, T., Kanei-Ishii, C., Maekawa, T., Fujimoto, J., Yamamoto, T., and Ishii, S. 1995. Increase in solubility of foreign proteins in *E. coli* by coproduction of the bacterial thioredoxin. *J. Biol. Chem.* 270:25328–25331.
- Zardeneta, G., and Horowitz, P. M. 1994. Detergent, liposome, and micelle-assisted protein refolding. *Anal. Biochem.* 223:1–6.
- Zhang, S. P., Zubay, G., and Goldman, E. 1991. Low-usage codons in *Escherichia coli*, yeast, fruit fly and primates. *Gene* 105:61–72.
- Zhi, W., Landry, S. J., Gierasch, L. M., and Srere, P. A. 1992. Renaturation of citrate synthase: Influence of denaturant and folding assistants. *Prot. Sci.* 1:522–529.